

1 **eRNA-IDO: A One-stop Platform for Identification, Interactome**  
2 **Discovery, and Functional Annotation of Enhancer RNAs**

3

4 Yuwei Zhang<sup>1,2,#</sup>, Lihai Gong<sup>3,#</sup>, Ruofan Ding<sup>3</sup>, Wenyan Chen<sup>3</sup>, Hao Rong<sup>4</sup>, Yanguo Li<sup>5</sup>,  
5 Fawziya Shameem<sup>1</sup>, Korakkandan Arshad Ali<sup>1</sup>, Lei Li<sup>3,\*</sup>, Qi Liao<sup>1,\*</sup>

6

7 <sup>1</sup> *School of Public Health, Health Science Center, Ningbo University, Ningbo 315211,*  
8 *China*

9 <sup>2</sup> *Biomedical Big Data Center, the First Affiliated Hospital, Zhejiang University School*  
10 *of Medicine, Hangzhou 310003, China*

11 <sup>3</sup> *Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen*  
12 *518107, China*

13 <sup>4</sup> *School of Clinical Medicine, Health Science Center, Ningbo University, Ningbo*  
14 *315211, China*

15 <sup>5</sup> *Institute of Drug Discovery Technology, Ningbo University, Ningbo 315211, China*

16

17 # Equal contribution.

18 \* Corresponding authors.

19 E-mail: liaoqi@nbu.edu.cn (Liao Q); lei.li@szbl.ac.cn (Li L).

20

21 **Running title:** *Zhang Y et al / eRNA-IDO Webserver*

22

23 Total word counts: 3062.

24 Total references: 54.

25 Total figures: 6.

26 Total tables: 1.

27 Total supplementary figures: 2.

28 Total supplementary tables: 8.

29 Total counts of letters in the article title: 102.

- 30 Total counts of letters in the running title: 28.
- 31 Total counts of keywords: 5.
- 32 Total words in abstract: 164.

### 33 **Abstract**

34 Growing evidence supports the transcription of enhancer RNAs (eRNAs) and their  
35 important roles in gene regulation. However, their interactions with other biomolecules  
36 and their corresponding functionality remain poorly understood. In an attempt to  
37 facilitate mechanistic research, this study presents eRNA-IDO, the first integrative  
38 computational platform for the identification, interactome discovery, and functional  
39 annotation of human eRNAs. eRNA-IDO comprises two modules: eRNA-ID and  
40 eRNA-Anno. Functionally, eRNA-ID can identify eRNAs from *de novo* assembled  
41 transcriptomes. eRNA-ID includes 8 kinds of enhancer makers, enabling users to  
42 customize enhancer regions flexibly and conveniently. In addition, eRNA-Anno  
43 provides cell-specific/tissue-specific functional annotation for both new and known  
44 eRNAs by analyzing the eRNA interactome from prebuilt or user-defined networks  
45 between eRNA and coding gene. The prebuilt networks include the Genotype-Tissue  
46 Expression (GTEx)-based co-expression networks in normal tissues, The Cancer  
47 Genome Atlas (TCGA)-based co-expression networks in cancer tissues, and omics-  
48 based eRNA-centric regulatory networks. eRNA-IDO can facilitate research on the  
49 biogenesis and functions of eRNAs. The eRNA-IDO server is freely available at  
50 [http://bioinfo.szbl.ac.cn/eRNA\\_IDO/](http://bioinfo.szbl.ac.cn/eRNA_IDO/).

51

52 **KEYWORDS:** Enhancer RNA; Identification; Interactome; Functional annotation;  
53 Webservice

## 54 **Introduction**

55 Over the past decade, a growing number of studies have reported the pervasive  
56 transcription of non-coding RNAs (ncRNAs) from active enhancer regions, termed  
57 enhancer RNAs (eRNAs). Due to the dynamic nature of enhancer activity across  
58 different tissues and lineages, eRNA transcription exhibits high specificity in biological  
59 contexts [1]. Once regarded as “transcription noise” or “byproduct” [2], eRNAs have  
60 now been shown to play crucial roles in various biological processes and diseases, such  
61 as cardiovascular development [3] and cancer [4]. Mechanistically, eRNAs can promote  
62 enhancer–promoter loops (E–P loops) and are involved in epigenetic regulation by  
63 interacting with other biomolecules, including components of cohesion or mediator  
64 [5,6], and histone acetyltransferases CBP/p300 [4,7]. Furthermore, eRNAs interact with  
65 transcription elongation factors to facilitate the pause-release of RNA polymerase II,  
66 thus controlling transcription elongation.

67 With the growing interest in eRNA functionality, several databases have been  
68 developed to characterize the transcription and potential targets of eRNAs, such as  
69 eRNAbase [8], Human enhancer RNA Atlas (HeRA) [9], the Cancer eRNA Atlas  
70 (TCeA) [10], Animal-eRNAdb [11], and eRNA in cancer (eRic) [12]. Nonetheless,  
71 these databases only provide information on annotated eRNA loci and enhancer regions,  
72 which do not allow the evaluation of novel eRNAs. Additionally, several platforms exist  
73 for functional annotation of ncRNAs, but they are not well-suited for eRNAs. For  
74 example, ncRNA functional annotation server (ncFANs) v2.0 [13] requires known  
75 ncRNA identifiers as input, but most eRNAs lack a reference ID or symbol. AnnoLnc2  
76 [14] allows the prediction of the functions of novel long ncRNAs (lncRNAs) based on  
77 co-expression networks but does not consider cell/tissue specificity and does not  
78 provide eRNA-specific characteristics such as histone modification, chromatin  
79 architecture, and interactive molecules. At present, a comprehensive platform for eRNA  
80 functional annotation is still lacking.

81 Therefore, this study introduces eRNA-IDO, the first one-stop platform for human  
82 eRNA identification, interactome discovery, and functional annotation (**Figure 1**).

83 eRNA-IDO comprises two available modules, namely eRNA-ID and eRNA-Anno.  
84 eRNA-ID enables users to define enhancers and identify enhancer-derived ncRNAs  
85 from uploaded *de novo* assembled transcriptome. eRNA-Anno predicts eRNA functions  
86 by discovering eRNA-connected protein-coding genes (PCGs) in normal/cancer co-  
87 expression and eRNA-centric regulatory networks. Furthermore, eRNA-IDO offers the  
88 capacity to utilize prebuilt data as well as user-defined data, providing a practical and  
89 convenient tool for biological researchers. This web server is freely available at  
90 [http://bioinfo.szbl.ac.cn/eRNA\\_IDO/](http://bioinfo.szbl.ac.cn/eRNA_IDO/) and is open to all users, without a login  
91 requirement.

92

## 93 **Method**

### 94 **Workflow and data architecture of eRNA-ID**

95 The left panel of Figure 1 illustrates the schematic workflow of eRNA-ID. The  
96 processing of *de novo* assembled transcripts is initiated from user-provided RNA  
97 sequencing (RNA-seq) or global run-on sequencing (GRO-seq) data. The transcripts  
98 overlapping with annotated PCGs, simple repeats, and blacklisted regions are removed  
99 based on the GENCODE v33 reference [15]. Thereafter, the coding potential of the  
100 remaining transcripts is evaluated by Coding Potential Calculator 2 (CPC2) [16]  
101 (default parameter), and ncRNAs transcribed from enhancer regions are identified as  
102 eRNAs. Enhancer regions can be either uploaded by users in Browser Extensible Data  
103 (BED) format or defined using our marker buffet. The marker buffet comprises 8 kinds  
104 of enhancer markers, including H3K27ac (Table S1), H3K4me1 (Table S2), chromatin  
105 accessibility (Table S3), RNA polymerase II binding (Table S4), super-enhancers from  
106 super-enhancer database (SEdb) 2.0 [17], EnhancerAtlas 2.0 [18] enhancers, functional  
107 annotation of the mammalian genome database (FANTOM5) [19] enhancers, and  
108 search candidate cis-regulatory elements by the encyclopedia of DNA elements  
109 (ENCODE) database (SCREEN) [20] enhancers. The markers are optionally  
110 overlapped or merged (using BEDTools multiinter/merge) to obtain high-confidence or  
111 comprehensive enhancer profiles. The  $\pm 3$  kb regions around the center of the selected

112 markers are defined as potential enhancer regions. These markers are cell-  
 113 specific/tissue-specific except those from FANTOM5 and SCREEN database. The data  
 114 type, source, and number of biosamples of these enhancer markers are listed in **Table**  
 115 **1**. Finally, eRNA-ID outputs the chromatin locations, adjacent genes (+/- 1Mb), and  
 116 enhancers of predicted eRNAs.

117

### 118 **Workflow and data architecture of eRNA-Anno**

119 The right panel of Figure 1 illustrates the schematic workflow of eRNA-Anno. The  
 120 chromatin coordinates of novel eRNAs in BED/gene transfer format (GTF) format or  
 121 the identifiers of known eRNAs annotated in HeRA [9] and eRic [12] databases are  
 122 input in eRNA-Anno. For known eRNAs, the ENSR identifiers, chromatin coordinates,  
 123 and adjacent genes (within +/- 1Mb) are accepted. Below is a detailed description of  
 124 each procedure.

125

#### 126 *eRNA quantification*

127 The expression levels of known eRNAs are obtained from HeRA and eRic. When  
 128 chromatin coordinates of novel eRNAs are input, RNA-seq data from TCGA  
 129 (<https://portal.gdc.cancer.gov/>) and GTEx portal [21] are used to quantify eRNA  
 130 expression. Subsequently, eRNA expression levels are estimated based on the read  
 131 coverage from BigWig files to shorten the processing time using the following formula:

132

$$FPKM = \frac{\sum(Cov) \times 10^9}{R \times L \times T}$$

133 Where  $\sum(Cov)$  represents the total read coverage of a given eRNA region,  $R$  is read  
 134 length,  $L$  is eRNA length, and  $T$  indicates the total mapped reads of the library.

135

#### 136 *Profiling genetic / epigenetic landscape*

137 eRNA-Anno portrays a genetic/epigenetic landscape for eRNAs, including chromatin  
 138 accessibility, clinically relevant mutation, and histone modification (H3K27ac and  
 139 H3K4me1). Histone modification and chromatin accessibility are characterized based  
 140 on chromatin immunoprecipitation sequencing (ChIP-seq) and assay for transposase-

141 accessible chromatin using sequencing (ATAC-seq)/DNase I hypersensitive sites  
142 sequencing (DNase-seq) from the Cistrome Data Browser [22] (Table S1–S3). Finally,  
143 clinically relevant mutations within the query eRNA regions are collected from ClinVar  
144 [23] and the Catalogue Of Somatic Mutations In Cancer (COSMIC) [24] database.

145

#### 146 *eRNA–PCG network construction*

147 Thereafter, a co-expression network between eRNAs and PCGs and an eRNA-centric  
148 regulatory network are constructed. The connected genes in the networks are defined  
149 as potential interactome of eRNAs. Both user-uploaded expression matrix and publicly  
150 available data are supported for the co-expression network. Publicly available data refer  
151 to RNA-seq data of 52 normal tissues from the GTEx portal [21] and 31 cancer types  
152 from the TCGA portal (Table S5). In addition, the toolkit GCEN [25] is used to calculate  
153 Spearman correlation coefficients and adjusted *P* values. The significant eRNA–PCG  
154 pairs are selected to construct the co-expression network according to user-defined  
155 thresholds.

156 For the eRNA-centric regulatory network, the relationships between eRNAs and  
157 transcription factor (TF), RNA binding protein (RBP), and E–P loop are analyzed. The  
158 eRNA–TF interactions are identified based on 11,356 ChIP-seq datasets from the  
159 Cistrome Data Browser [22], which involve 1354 TFs and 642 cells/tissues (Table S4).  
160 Furthermore, the eRNA–RBP interactions are obtained based on 518 cross-linking  
161 immunoprecipitation sequencing (CLIP-seq) datasets from the post-transcriptional  
162 regulation coordinated by RBP (POSTAR3) database [26], which involve 221 RBPs  
163 and 34 cells/tissues (Table S6). TFs and RBPs with peaks located within eRNA regions  
164 are defined as potential regulators of eRNAs. E–P loops identified by 198 HiChIP  
165 experiments across 108 cell types (Table S7) are collected from HiChIPdb [27]. The  
166 loops harboring anchors overlapping with query eRNAs are defined as eRNA-mediated  
167 loops.

168

#### 169 *Subnetwork extraction*

170 Subsequently, eRNA-Anno extracts hubs/modules from the overall network to obtain

171 the tightly connected PCGs of query eRNAs. During this process, SPICi [28] in the  
172 unweighted mode (default parameter) is utilized for module extraction.

173

### 174 *Functional enrichment analyses*

175 Functional enrichment analyses, including Gene Ontology (GO), Kyoto Encyclopedia  
176 of Genes and Genomes (KEGG) pathway, and Molecular Signatures Database  
177 (MSigDB) hallmark enrichment [29], are performed based on hypergeometric tests  
178 using our in-house scripts (<https://github.com/zhangyw0713/FunctionEnrichment>).

179

## 180 **Results**

### 181 **Web interface of eRNA-ID**

182 eRNA-ID has been designed for eRNA identification based on *de novo* assembled  
183 transcriptome. In the input interface ([http://bioinfo.szbl.ac.cn/eRNA\\_IDO/eRNA-ID](http://bioinfo.szbl.ac.cn/eRNA_IDO/eRNA-ID)),  
184 users are required to upload a transcriptome profile in GTF format, which can be  
185 generated from RNA-seq and GRO-seq data, and define enhancer regions using our  
186 marker buffet or by uploading their BED file. eRNA-ID adopts a similar analytical  
187 workflow to the one used in ncFANs-eLnc [13] to identify eRNAs (see Method). As  
188 shown in Table S8, the major advantage of eRNA-ID compared to ncFANs is the  
189 inclusion of a prebuilt buffet of 8 kinds of enhancer markers (H3K27ac, H3K4me1,  
190 chromatin accessibility, RNA polymerase II binding, SEdb 2.0 super-enhancers [17],  
191 and three types of enhancer annotations from EnhancerAtlas 2.0 [18], FANTOM5 [19],  
192 and SCREEN [20] databases), enabling users to customize enhancer regions of interest.  
193 For example, users may require high-confidence enhancer regions simultaneously  
194 labeled by multiple markers or may want to obtain as many enhancers as possible by  
195 merging all markers. The processing procedure of eRNA-ID is fast; a GRO-seq-derived  
196 transcriptome with 3483 transcripts (SRA008244) took 45 seconds, and a total RNA-  
197 seq-derived *de novo* transcriptome with 222,848 transcripts (GSM2824220) took 88  
198 seconds (default parameters).

199 In the output interface of eRNA-ID (

200 **Figure 2**), the chromatin coordinates, enhancers, and putative targets (adjacent  
201 genes within  $\pm$  1Mb of eRNAs) of identified eRNAs are displayed in a table. Users  
202 can also view the information in a genome browser based on JBrowse [30]. Moreover,  
203 functional annotation can be conducted for these novel eRNAs by clicking on the  
204 “Deliver eRNA to eRNA-Anno” button.

205

### 206 **Web interface of eRNA-Anno**

207 eRNA-Anno has been designed for the network-based interactome discovery and  
208 functional annotation of eRNAs. In this module, users input the chromatin coordinates  
209 of novel eRNAs (**Figure 3A**) or the identifiers/locations of known eRNAs annotated in  
210 HeRA [9] and eRic [12] databases (Figure 3B), followed by network selection and  
211 parameter setting. eRNA-Anno first quantifies the eRNA expression levels based on  
212 RNA-seq data from TCGA and GTEx portal. As hundreds of RNA-seq samples require  
213 a long processing time, the read coverages from BigWig files were used to speed up the  
214 quantification (see Method). To examine the reliability of this method, the expression  
215 levels of known eRNAs acquired via this method were correlated with those based on  
216 the canonical featureCounts [31] method obtained from HeRA and eRic databases. The  
217 results revealed that our method was highly correlated with the canonical method  
218 (Figure S1A and B) and is approximately 400 times faster (Figure S1C).

219 Next, eRNA-IDO is used to annotate the functions of eRNAs by discovering their  
220 interactomes. Interactome discovery is based on eRNA-PCG networks, including  
221 normal co-expression networks based on GTEx expression profiles [21], cancer co-  
222 expression networks based on TCGA expression profiles  
223 (<https://portal.gdc.cancer.gov/>), and eRNA-centric regulatory networks. Co-expression  
224 relationships are widely used to annotate the functions of eRNAs [32–34]. Additionally,  
225 eRNAs were reported to exert regulatory functions by interacting with other  
226 biomolecules, such as TFs [35–37], RBPs [4,38,39], and target genes activated by E-P  
227 loops [40,41]. Therefore, the regulatory network can be used for eRNA functional  
228 annotation, resembling those used for other ncRNAs [13,42–45]. The network  
229 construction procedure is detailed in the Method section. Parameters include

230 tissue/cancer type of expression profile, co-expression coefficient, significance  
231 threshold, biosamples of interaction relationships, and epigenetic landscape (Figure 3C  
232 and D).

233 Upon receiving launch instructions, eRNA-Anno initiates the analytical procedure  
234 (see Method) to identify the potential targets of query eRNAs from the selected  
235 networks and annotate their functions based on hub-based and module-based strategies.  
236 The whole procedure typically takes tens of minutes, depending on the number of input  
237 eRNAs (Figure S2). Hence, users are recommended to set an email notification or  
238 record the task ID for result retrieval when submitting a task with a large set of eRNAs.

239 In the output interface, eRNA-Anno provides basic information about eRNAs (*i.e.*,  
240 location and expression, epigenetic landscape, and disease relevance) and putative  
241 targets and functions based on the various networks. In the “Location and expression”  
242 section, chromatin coordinates, the expression level in normal and cancer samples,  
243 adjacent genes ( $\leq 1$  Mb), and overlapped super-enhancers are listed in the table  
244 (**Figure 4A**). Furthermore, eRNA-Anno profiles active enhancer markers (H3K27ac  
245 and H3K4me1) and chromatin accessibility of eRNA regions to evaluate the activity of  
246 enhancers where eRNAs are transcribed (Figure 4B). Considering that mutations in  
247 eRNA regions are often related to eRNA expression and subsequent disease  
248 development [46], clinically relevant mutations within query eRNA regions are  
249 displayed in the “Disease relevance” section (Figure 4C) and can be visualized in  
250 genome browser (Figure 4D). Moreover, the interactome and predicted functions of  
251 eRNAs based on the selected networks are displayed in the second part (**Figure 5**). For  
252 example, in a cancer co-expression network (Figure 5A), the eRNA–PCG network is  
253 visualized in a force-directed layout, and the functions of connected PCGs are provided  
254 (Figure 5B). Since genes with similar functions tend to be concentrically distributed,  
255 eRNA-Anno then extracts hubs and modules composed of tightly connected genes from  
256 the overall network (Figure 5C). The function of query eRNAs can be inferred by the  
257 functions of the PCGs within the same module or hub (Figure 5D).

258 In addition, the eRNA-centric regulatory network (Figure 5E) provides a  
259 visualization of the relationships of eRNAs with TFs, RBPs, and E–P loops in multiple

260 modes, including network topology, table, and genome browser. Similarly, the functions  
261 of eRNAs can be inferred by the related biomolecules in the overall network, modules,  
262 or hubs. The results of individual networks can be combined into a summary (**Figure**  
263 **6**).

264

### 265 **A case study demonstrating the usage of eRNA-Anno**

266 Since the input interface has many user-dependent options and the output interface  
267 displays interactive information, a case study is described to introduce the usage and  
268 interpretation of results obtained from eRNA-Anno. CCAT1 and LINC02257, which  
269 have been characterized as colon cancer-associated eRNAs [47,48], were analyzed in  
270 this study and input in GTF format. Next, “TCGA-COAD” and “GTEx-Colon  
271 Transverse” were chosen, co-expression and regulatory networks were selected, the  
272 parameters were set, and eRNA-IDO was finally launched, as depicted in Figure 3.

273 In the output interface, eRNA-Anno revealed that both CCAT1 and LINC02257  
274 exhibited higher expression levels in colorectal cancer (Figure 4A) and showed  
275 enriched active enhancer markers (Figure 4B), which was consistent with previously  
276 published studies [47,48]. Additionally, the genomic regions of CCAT1 and  
277 LINC02257 harbor carcinoma-associated mutations (Figure 4C), indicating their  
278 clinical significance. Subsequently, the co-expression network in colon  
279 adenocarcinoma was further investigated to evaluate the interactome and functions of  
280 CCAT1 and LINC02257. The topology of the co-expression network revealed limited  
281 connections between CCAT1 and LINC02257 (Figure 5A), indicating their independent  
282 regulatory roles. Furthermore, functional enrichment analysis of the co-expressed  
283 PCGs revealed that CCAT1 and LINC02257 were potentially enriched in translation  
284 and cell cycle pathways (Figure 5B). The module involved in CCAT1 precisely  
285 pinpointed the role of CCAT1 in regulating the cell cycle (Figure 5C and D), which  
286 conforms to previous findings [30,49]. Moreover, the eRNA-centric regulatory network  
287 detected the interactive TFs, RBPs, and genes involved in E–P loops. These interactive  
288 molecules were enriched in cell cycle and cancer pathways, suggesting the similar role  
289 of CCAT1 and LINC02257 (Figure 5E). Additionally, a genome browser based on

290 JBrowse [30] was provided to visualize eRNA locations and the mutational, epigenetic,  
291 and interactive landscapes (Figure 5E). Finally, the nodes and edges from two separate  
292 networks were overlapped to determine high-confidence interactions of CCAT1 in a  
293 cell cycle-related module (Figure 6); some targets such as CDK4 [50] and SOX4 [51]  
294 had been previously reported. This case study demonstrates the application of eRNA-  
295 Anno, displaying its ability to comprehensively and reliably predict the eRNA  
296 interactome and functions.

297

## 298 Discussion

299 As a web server dedicated to eRNA analysis, eRNA-IDO provides a convenient method  
300 of eRNA identification, interactome discovery, and functional annotation. The major  
301 advantages of eRNA-IDO include but are not limited to the following. First, eRNA-ID  
302 includes 8 kinds of enhancer markers, offering a more convenient and customized  
303 approach for enhancer definition compared to ncFANs-eLnc [13], which only includes  
304 the H3K27ac marker. Second, eRNA-Anno is applicable to both novel and known  
305 eRNAs. Considering the poor characterization of eRNAs, the applicability to novel  
306 eRNAs grants eRNA-Anno higher flexibility and biological practicability compared to  
307 other tools requiring known identifiers such as ncFANs v2.0 [13] and other databases  
308 [9–12]. The detailed comparison between eRNA-IDO and ncFANs v2.0 is displayed in  
309 Table S8. Third, biological context-specific expression and interaction profiles are  
310 prebuilt in eRNA-Anno. Compared to tools without biological specificity such as  
311 AnnoLnc2 [14], eRNA-Anno is expected to provide more precise findings for *in vivo*  
312 investigations. Moreover, the prebuilt profiles facilitate the use of the service. Finally,  
313 eRNA-IDO is the first one-stop platform for eRNA identification, interactome  
314 discovery, and functional annotation.

315 Nevertheless, the limitations of the study should be acknowledged and overcome.  
316 First, eRNA-IDO is currently designed for human data and additional species will be  
317 supported in the future. Second, some characteristics such as the m<sup>6</sup>A modification [52]  
318 and RNA structure [53,54] are essential for eRNA functionality but are not evaluated

319 by eRNA-IDO. Third, the current iteration of eRNA-IDO only considers normal tissue  
320 and cancer. In the future, a larger number of disease-specific and cell-specific  
321 expression and interaction profiles will be incorporated. Hopefully, eRNA-IDO will  
322 benefit from user feedback and develop into a more powerful tool upon continuous  
323 updates.

324

### 325 **Data availability**

326 The eRNA-IDO webserver is available at [http://bioinfo.szbl.ac.cn/eRNA\\_IDO/](http://bioinfo.szbl.ac.cn/eRNA_IDO/).

327

### 328 **CRedit author statement**

329 **Yuwei Zhang:** Data curation, Methodology, Investigation, Writing - original draft.

330 **Lihai Gong:** Methodology, Software, Visualization. **Ruofan Ding:** Data curation,

331 Visualization. **Wenyan Chen:** Data curation, Investigation. **Hao Rong:** Data curation,

332 Investigation. **Yanguo Li:** Data curation. **Fawziya Shameem:** Writing - review &

333 editing. **Korakkandan Arshad Ali:** Writing - review & editing. **Lei Li:** Resources,

334 Conceptualization, Supervision, Writing - review & editing. **Qi Liao:**

335 Conceptualization, Supervision, Writing - review & editing, Funding acquisition. All

336 authors have read and approved the final manuscript.

337

### 338 **Competing interests**

339 The authors have declared no competing interests.

340

### 341 **Supplementary material**

342 Supplementary material is available at *Genomics, Proteomics & Bioinformatics* online

343 (<https://doi.org/10.1093/gpbjnl/qzaxxxx>).

344

### 345 **Acknowledgments**

346 This work was supported by the National Natural Science Foundation of China (Grant  
347 Nos. 32300430, 32100533, and 31970630), Open Grant Funds from Shenzhen Bay  
348 Laboratory to Lei Li (Grant No. SZBL2021080601001), the Zhejiang Provincial  
349 Natural Science Foundation of (Grant No. LY21C060002), and the Ningbo Major  
350 Research and Development Plan Project (Grant Nos. 2023Z226 and 2023Z171).

351

## 352 **ORCID**

353 0000-0002-8395-3020 (Yuwei Zhang)

354 0009-0007-0880-8177 (Lihai Gong)

355 0000-0003-1667-9866 (Ruofan Ding)

356 0000-0002-4821-9306 (Wenyan Chen)

357 0000-0002-1650-0670 (Hao Rong)

358 0000-0001-9064-0033 (Yanguo Li)

359 0009-0009-5011-7665 (Fawziya Shameem)

360 0009-0002-9962-0543 (Korakkandan Arshad Ali)

361 0000-0003-3924-2544 (Lei Li)

362 0000-0001-6796-104X (Qi Liao)

363

## 364 **References**

365 [1] Lam MT, Li W, Rosenfeld MG, Glass CK. Enhancer RNAs and regulated  
366 transcriptional programs. *Trends Biochem Sci* 2014;39:170–82.

367 [2] Han Z, Li W. Enhancer RNA: what we know and what we can achieve. *Cell Prolif*  
368 2022;55:e13202.

369 [3] Ounzain S, Pedrazzini T. Super-enhancer lncs to cardiovascular development and  
370 disease. *Biochim Biophys Acta* 2016;1863:1953–60.

371 [4] Jiao W, Chen Y, Song H, Li D, Mei H, Yang F, et al. HPSE enhancer RNA promotes  
372 cancer progression through driving chromatin looping and regulating  
373 hnRNPU/p300/EGR1/HPSE axis. *Oncogene* 2018;37:2728–45.

374 [5] Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, et al. Activating  
375 RNAs associate with Mediator to enhance chromatin architecture and transcription.  
376 *Nature* 2013;494:497–501.

- 377 [6] Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, et al. Functional roles of  
378 enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*  
379 2013;498:516–20.
- 380 [7] Bose DA, Donahue G, Reinberg D, Shiekhattar R, Bonasio R, Berger SL. RNA  
381 binding to CBP stimulates histone acetylation and transcription. *Cell* 2017;168:135–  
382 49.e22.
- 383 [8] Song C, Zhang G, Mu X, Feng C, Zhang Q, Song S, et al. eRNAbase: a  
384 comprehensive database for decoding the regulatory eRNAs in human and mouse.  
385 *Nucleic Acids Res* 2024;52:D81–91.
- 386 [9] Zhang Z, Hong W, Ruan H, Jing Y, Li S, Liu Y, et al. HeRA: an atlas of enhancer  
387 RNAs across human tissues. *Nucleic Acids Res* 2021;49:D932–8.
- 388 [10] Chen H, Liang H. A high-resolution map of human enhancer RNA loci  
389 characterizes super-enhancer activities in cancer. *Cancer Cell* 2020;38:701–15.e5.
- 390 [11] Jin W, Jiang G, Yang Y, Yang J, Yang W, Wang D, et al. Animal-eRNAdb: a  
391 comprehensive animal enhancer RNA database. *Nucleic Acids Res* 2022;50:D46–53.
- 392 [12] Zhang Z, Lee JH, Ruan H, Ye Y, Krakowiak J, Hu Q, et al. Transcriptional  
393 landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer.  
394 *Nat Commun* 2019;10:4562.
- 395 [13] Zhang Y, Bu D, Huo P, Wang Z, Rong H, Li Y, et al. ncFANs v2.0: an integrative  
396 platform for functional annotation of non-coding RNAs. *Nucleic Acids Res*  
397 2021;49:W459–68.
- 398 [14] Ke L, Yang DC, Wang Y, Ding Y, Gao G. AnnoLnc2: the one-stop portal to  
399 systematically annotate novel lncRNAs for human and mouse. *Nucleic Acids Res*  
400 2020;48:W230–8.
- 401 [15] Frankish A, Diekhans M, Jungreis I, Lagarde J, Loveland JE, Mudge JM, et al.  
402 GENCODE 2021. *Nucleic Acids Res* 2021;49:D916–23.
- 403 [16] Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, et al. CPC2: a fast and  
404 accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids*  
405 *Res* 2017;45:W12–6.
- 406 [17] Wang Y, Song C, Zhao J, Zhang Y, Zhao X, Feng C, et al. SEdb 2.0: a  
407 comprehensive super-enhancer database of human and mouse. *Nucleic Acids Res*  
408 2023;51:D280–90.
- 409 [18] Gao T, Qian J. EnhancerAtlas 2.0: an updated resource with enhancer annotation  
410 in 586 tissue/cell types across nine species. *Nucleic Acids Res* 2020;48:D58–64.
- 411 [19] Abugessaisa I, Ramiłowski JA, Lizio M, Severin J, Hasegawa A, Harshbarger J, et  
412 al. FANTOM enters 20th year: expansion of transcriptomic atlases and functional  
413 annotation of non-coding RNAs. *Nucleic Acids Res* 2021;49:D892–8.

- 414 [20] The ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB,  
415 Shosh N, et al. Expanded encyclopaedias of DNA elements in the human and mouse  
416 genomes. *Nature* 2020;583:699–710.
- 417 [21] GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects  
418 across human tissues. *Science* 2020;369:1318–30.
- 419 [22] Zheng R, Wan C, Mei S, Qin Q, Wu Q, Sun H, et al. Cistrome Data Browser:  
420 expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res*  
421 2019;47:D729–35.
- 422 [23] Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar:  
423 improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*  
424 2018;46:D1062–7.
- 425 [24] Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC:  
426 the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* 2019;47:D941–7.
- 427 [25] Chen W, Li J, Huang S, Li X, Zhang X, Hu X, et al. GCEN: an easy-to-use toolkit  
428 for gene co-expression network analysis and lncRNAs annotation. *Curr Issues Mol Biol*  
429 2022;44:1479–87.
- 430 [26] Zhao W, Zhang S, Zhu Y, Xi X, Bao P, Ma Z, et al. POSTAR3: an updated platform  
431 for exploring post-transcriptional regulation coordinated by RNA-binding proteins.  
432 *Nucleic Acids Res* 2022;50:D287–94.
- 433 [27] Wang S, Gao S, Zeng Y, Zhu L, Mo Y, Wong CC, et al. *N*<sup>6</sup>-methyladenosine reader  
434 YTHDF1 promotes ARHGEF2 translation and RhoA signaling in colorectal cancer.  
435 *Gastroenterology* 2022;162:1183–96.
- 436 [28] Jiang P, Singh M. SPICi: a fast clustering algorithm for large biological networks.  
437 *Bioinformatics* 2010;26:1105–11.
- 438 [29] Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The  
439 Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*  
440 2015;1:417–25.
- 441 [30] Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, Leung A, et al.  
442 JBrowse 2: a modular genome browser with views of synteny and structural variation.  
443 *Genome Biol* 2023;24:74.
- 444 [31] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for  
445 assigning sequence reads to genomic features. *Bioinformatics* 2014;30:923–30.
- 446 [32] Cai H, Liang J, Jiang Y, Tan R, Hou C, Hou J. Integrative analysis of *N*<sup>6</sup>-  
447 methyladenosine-related enhancer RNAs identifies distinct prognosis and tumor  
448 immune micro-environment patterns in head and neck squamous cell carcinoma.  
449 *Cancers (Basel)* 2022;14:4657.
- 450 [33] Chen X, Yuan J, Xue G, Campanario S, Wang D, Wang W, et al. Translational  
451 control by DHX36 binding to 5'UTR G-quadruplex is essential for muscle stem-cell

- 452 regenerative functions. *Nat Commun* 2021;12:5043.
- 453 [34] Yao P, Lin P, Gokoolparsadh A, Assareh A, Thang MW, Voineagu I. Coexpression  
454 networks identify brain region-specific enhancer RNAs in the human brain. *Nat*  
455 *Neurosci* 2015;18:1168–74.
- 456 [35] Allen MA, Andrysiak Z, Dengler VL, Mellert HS, Guarnieri A, Freeman JA, et al.  
457 Global analysis of p53-regulated transcription identifies its direct targets and  
458 unexpected regulatory mechanisms. *Elife* 2014;3:e02200.
- 459 [36] Azofeifa JG, Allen MA, Hendrix JR, Read T, Rubin JD, Dowell RD. Enhancer  
460 RNA profiling predicts transcription factor activity. *Genome Res* 2018;28:334–44.
- 461 [37] Franco HL, Nagari A, Malladi VS, Li W, Xi Y, Richardson D, et al. Enhancer  
462 transcription reveals subtype-specific gene expression programs controlling breast  
463 cancer pathogenesis. *Genome Res* 2018;28:159–70.
- 464 [38] Bai X, Li F, Zhang Z. A hypothetical model of *trans*-acting R-loops-mediated  
465 promoter–enhancer interactions by Alu elements. *J Genet Genomics* 2021;48:1007–19.
- 466 [39] Huang Z, Yu H, Du G, Han L, Huang X, Wu D, et al. Enhancer RNA lnc-CES1-1  
467 inhibits decidual cell migration by interacting with RNA-binding protein FUS and  
468 activating PPARgamma in URPL. *Mol Ther Nucleic Acids* 2021;24:104–12.
- 469 [40] Arnold PR, Wells AD, Li XC. Diversity and emerging roles of enhancer RNA in  
470 regulation of gene expression and cell fate. *Front Cell Dev Biol* 2020;7:377.
- 471 [41] Harrison LJ, Bose D. Enhancer RNAs step forward: new insights into enhancer  
472 function. *Development* 2022;149:dev200398.
- 473 [42] Zhang Y, Tao Y, Li Y, Zhao J, Zhang L, Zhang X, et al. The regulatory network  
474 analysis of long noncoding RNAs in human colorectal cancer. *Funct Integr Genomics*  
475 2018;18:261–75.
- 476 [43] Zhang Y, Tao Y, Liao Q. Long noncoding RNA: a crosslink in biological regulatory  
477 network. *Brief Bioinform* 2018;19:930–45.
- 478 [44] Chen L, Zhang W, Li DY, Wang X, Tao Y, Zhang Y, et al. Regulatory network  
479 analysis of LINC00472, a long noncoding RNA downregulated by DNA  
480 hypermethylation in colorectal cancer. *Clin Genet* 2018;93:1189–98.
- 481 [45] Luo C, Tao Y, Zhang Y, Zhu Y, Minyao DN, Haleem M, et al. Regulatory network  
482 analysis of high expressed long non-coding RNA LINC00941 in gastric cancer. *Gene*  
483 2018;662:103–9.
- 484 [46] Hu X, Wu L, Yao Y, Ma J, Li X, Shen H, et al. The integrated landscape of eRNA  
485 in gastric cancer reveals distinct immune subtypes with prognostic and therapeutic  
486 relevance. *iScience* 2022;25:105075.
- 487 [47] McClelland ML, Mesh K, Lorenzana E, Chopra VS, Segal E, Watanabe C, et al.  
488 CCAT1 is an enhancer-templated RNA that predicts BET sensitivity in colorectal cancer.

- 489 J Clin Invest 2016;126:639–52.
- 490 [48] Xiao J, Liu Y, Yi J, Liu X. LINC02257, an enhancer RNA of prognostic value in  
491 colon adenocarcinoma, correlates with multi-omics immunotherapy-related analysis in  
492 33 cancers. *Front Mol Biosci* 2021;8:646786.
- 493 [49] Liu Z, Chen Q, Hann SS. The functions and oncogenic roles of CCAT1 in human  
494 cancer. *Biomed Pharmacother* 2019;115:108943.
- 495 [50] Li JL, Li R, Gao Y, Guo WC, Shi PX, Li M. LncRNA CCAT1 promotes the  
496 progression of preeclampsia by regulating CDK4. *Eur Rev Med Pharmacol Sci*  
497 2018;22:1216–23.
- 498 [51] Hu B, Zhang H, Wang Z, Zhang F, Wei H, Li L. LncRNA CCAT1/miR-130a-3p  
499 axis increases cisplatin resistance in non-small-cell lung cancer cell line by targeting  
500 SOX4. *Cancer Biol Ther* 2017;18:974–83.
- 501 [52] Lee JH, Wang R, Xiong F, Krakowiak J, Liao Z, Nguyen PT, et al. Enhancer RNA  
502 m<sup>6</sup>A methylation facilitates transcriptional condensate formation and gene activation.  
503 *Mol Cell* 2021;81:3368–85.e9.
- 504 [53] Cheng JH, Pan DZ, Tsai ZT, Tsai HK. Genome-wide analysis of enhancer RNA in  
505 gene regulation across 12 mouse tissues. *Sci Rep* 2015;5:12648.
- 506 [54] Ren C, Liu F, Ouyang Z, An G, Zhao C, Shuai J, et al. Functional annotation of  
507 structural ncRNAs within enhancer RNAs in the human genome: implications for  
508 human disease. *Sci Rep* 2017;7:15518.

## 509 **Figure legends**

### 510 **Figure 1 The workflow of eRNA-IDO**

511 eRNA-IDO comprises two functional modules, eRNA-ID for eRNA identification and  
512 eRNA-Anno for interactome discovery and functional annotation. eRNA, enhancer  
513 RNA; SEdb, super-enhancer database; FANTOM5, functional annotation of the  
514 mammalian genome database; ENCODE, the encyclopedia of DNA elements;  
515 SCREEN, search candidate cis-regulatory elements by ENCODE; RNAP II, RNA  
516 polymerase II; KEGG, kyoto encyclopedia of genes and genomes; MSigDB, the  
517 molecular signatures database; RBP, RNA binding protein; TSS, transcription start site;  
518 TF, transcription factor; BED, browser extensible data; GTF, gene transfer format;  
519 CPC2, coding potential calculator 2.

520

### 521 **Figure 2 The output interface of eRNA-ID**

522 The predicted eRNA locations, enhancer regions, markers for active enhancers, putative  
523 targets (adjacent genes), and overlapped lncRNAs are displayed in a table and can be  
524 visualized in the genome browser. Additional details are shown in the demo:  
525 [http://bioinfo.szbl.ac.cn/eRNA\\_IDO/retrieve/?taskid=5a9LFXS8oGCm](http://bioinfo.szbl.ac.cn/eRNA_IDO/retrieve/?taskid=5a9LFXS8oGCm). lncRNAs,  
526 long non-coding RNAs.

527

### 528 **Figure 3 The input interface of eRNA-Anno**

529 **A.** The input contents include a potential eRNA list, optional target candidates,  
530 parameters for eRNA quantification, network selection, and genetic/epigenetic  
531 landscape. **B.** The input interface for known eRNAs annotated in HeRA [9] and eRic  
532 [12]. **C.** Parameters for the construction of the co-expression network. **D.** Parameters  
533 for the construction of the eRNA-centric regulatory network. HeRA, Human enhancer  
534 RNA Atlas; eRic, eRNA in cancer.

535

536 **Figure 4 The output interface of eRNA-Anno shows the basic information of**  
537 **query eRNA CCAT1 and LINC02257**

538 **A.** The locations and the expression levels of CCAT1 and LINC02257. **B.** The  
 539 epigenetic landscape. **C.** Clinically relevant mutations within the genomic regions of  
 540 CCAT1 and LINC02257. **D.** The genome browser can be activated by clicking on the  
 541 button “Visualization in genome browser”. Further details are shown in the demo:  
 542 [http://bioinfo.szbl.ac.cn/eRNA\\_IDO/retrieve/?taskid=97XPLicEAj4euYG/](http://bioinfo.szbl.ac.cn/eRNA_IDO/retrieve/?taskid=97XPLicEAj4euYG/).

543

544 **Figure 5 The output interface of eRNA-Anno shows the interactomes and**  
 545 **functions of CCAT1 and LINC02257**

546 **A.** The co-expression network of CCAT1 and LINC02257 in human colorectal cancer.  
 547 **B.** The enriched KEGG pathways of CCAT1- and LINC02257-connected PCGs. **C.**  
 548 Visualization of the LINC02257-containing module. **D.** The enriched KEGG pathways  
 549 of the PCGs within the LINC02257-containing module. **E.** The CCAT1-centric and  
 550 LINC02257-centric regulatory network.

551

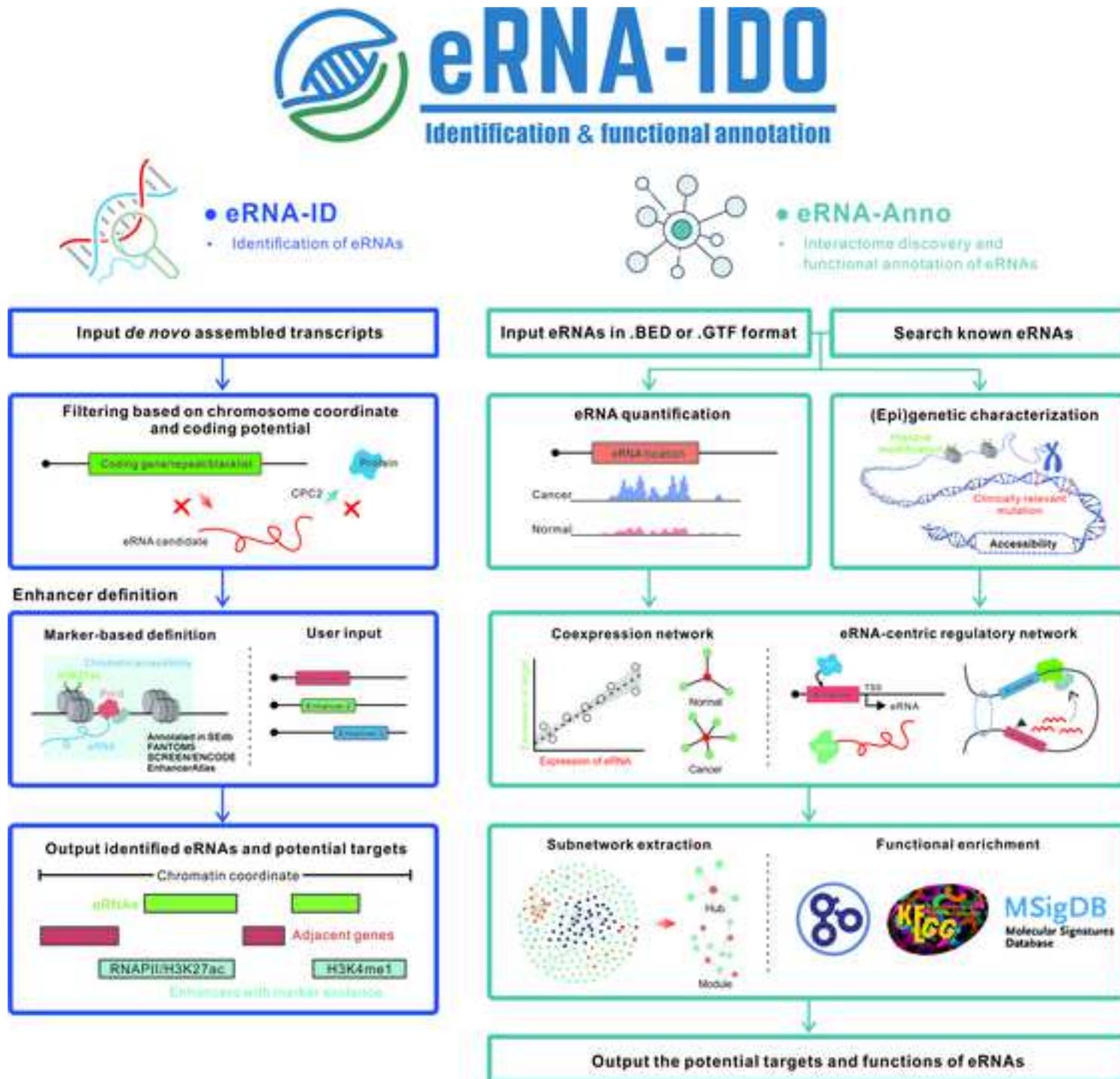
552 **Figure 6 Summary of the interactome and functions of query eRNAs based on**  
 553 **the combination of co-expression network and regulatory network**

554 **A.** Parameter settings for network combination. **B.** A high-confidence network  
 555 comprising the overlapped nodes and edges was generated for CCAT1 and LINC02257.  
 556 **C.** The module involved in CCAT1 indicates its interactive genes and functions in cell  
 557 cycle regulation.

558

559 **Table**

560 **Table 1 Data type, source, and the number of biosamples of enhancer markers**



Home / eRNA\_ID

Enhancer count: 137084

Novel ncRNA count: 512

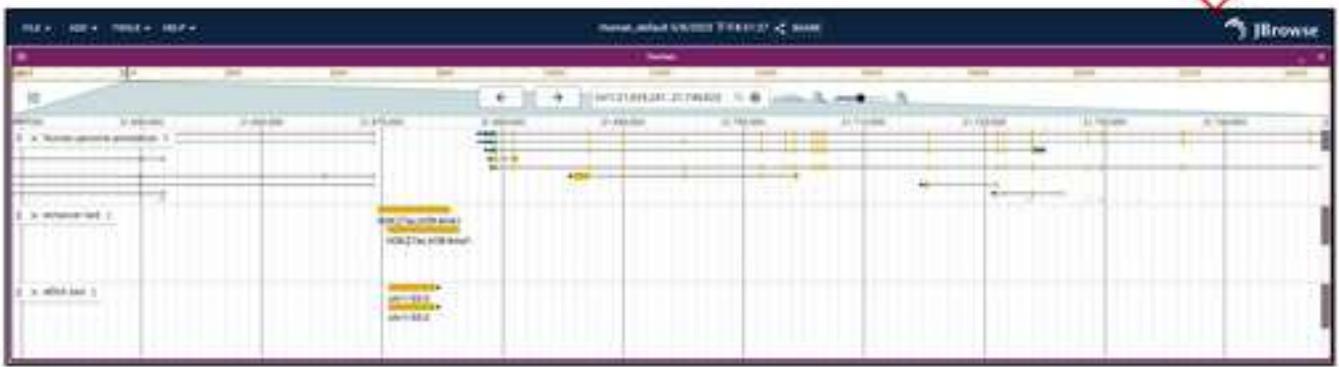
eRNA (Novel ncRNA from enhancer) count: 389

[Download the eRNA list](#) [Download CDS report](#)

eRNA_id	eRNA_start	eRNA_end	eRNA_ID	Strand	Enhancer_chr	Enhancer_start	Enhancer_end	Enhancer_marks	Pubmed targets	Outstanding
Chr1	11252163	11254317	chr1-26-0	-	Chr1	11262014	11262015	H3K9me1	<a href="#">View positive targets</a>	<a href="#">View</a>
Chr1	16076019	16078472	chr1-39-0	+	Chr1	16677160	16681591	H3K9me1	<a href="#">View positive targets</a>	<a href="#">View</a>
Chr1	21670040	21674360	chr1-55-0	-	Chr1	21609662	21670965	H3K27ac,H3K9me1	<a href="#">View positive targets</a>	<a href="#">View</a>
Chr1	21670040	21674360	chr1-55-0	+	Chr1	21670438	21674419	H3K27ac,H3K9me1	<a href="#">View positive targets</a>	<a href="#">View</a>
Chr1	32064797	32067711	chr1-423-1	-	Chr1	32061549	32061549	H3K9me1	<a href="#">View positive targets</a>	<a href="#">View</a>
Chr1	32064797	32067711	chr1-423-1	-	Chr1	32060949	32061947	H3K27ac,H3K9me1	<a href="#">View positive targets</a>	<a href="#">View</a>
Chr1	32064797	32067711	chr1-423-1	-	Chr1	32061331	32067531	H3K27ac	<a href="#">View positive targets</a>	<a href="#">View</a>
Chr1	36386119	36389968	chr1-401-1	-	Chr1	36380158	36380158	H3K27ac,H3K9me1	<a href="#">View positive targets</a>	<a href="#">View</a>
Chr1	36453201	36456426	chr1-127-0	+	Chr1	36448025	36454025	H3K27ac,H3K9me1	<a href="#">View positive targets</a>	<a href="#">View</a>
Chr1	42172194	42175574	chr1-190-0	+	Chr1	42171619	42175719	H3K9me1	<a href="#">View positive targets</a>	<a href="#">View</a>

813 items | 1 2 3 4 5 ... 82 83 101 Done

[Delete eRNA to eRNA-Atlas](#) [Visualization in genome browser](#)



**A**

Novel sRNA **Known sRNA** **Move to input page for knowRNA**

User input

1. Upload a query sRNA for J+3256 is recommended. [Test example](#)

2. Input candidate target genes to constrain the network (Optional). [Test example](#)

Expression quantification

Please define expression profiles for sRNA quantification and an expression network association:

Based on the preconstructed expression profiles:

Disease: 
  
 Normal:

Or users upload the expression profiles (both below are required). [Test example](#)

Network Selection

Normal and disease co-expression network
   
 Regulatory relationships

Genetic/Epigentic landscape

Epigenetic feature
   
 Mutation and Disease relevance

**Run**

**B**

Novel sRNA **Known sRNA**

User input

These query sRNAs for J+3256 are recommended. [Test example](#) Or upload the file via the bottom.

Upload

**C**

Network Selection

Normal and disease co-expression network

NOTE: co-expression network construction is based on the expression profiles you selected / uploaded in above "Expression quantification" section.

Allow to adjust quantitative parameters

Network association coefficient between sRNA and protein-coding gene

0.4

Network association coefficient between sRNA and protein-coding gene

0.1

Equilibrium threshold

0.05

Select the regulatory relationships

Regulatory relationships

**D**

Regulatory relationships

Expression factor

sRNA binding protein

Evidence promoter loop

**Choose biosample**







**Table 1 Data type, source, and the number of biosamples of enhancer markers**

<b>Marker</b>	<b>Data type</b>	<b>Data source</b>	<b>Biosample</b>	<b>Ref.</b>
Chromatin accessibility	ATAC-seq/DNase-seq	Cistrome	371	[47]
H3K27ac	ChIP-seq	Cistrome	555	[47]
H3K4me1	ChIP-seq	Cistrome	364	[47]
Polymerase II	ChIP-seq	Cistrome	166	[47]
FANTOM5 enhancer	-	FANTOM5	-	[17]
SCREEN enhancer	-	SCREEN	-	[18]
EnhancerAtlas enhancer	-	EnhancerAtlas 2.0	197	[16]
Super-enhancer	-	SEdb 2.0	1705	[46]

*Note:* ATAC-seq, assay for transposase-accessible chromatin using sequencing; DNase-seq, DNase I hypersensitive sites sequencing; ChIP-seq, chromatin immunoprecipitation sequencing; FANTOM5, functional annotation of the mammalian genome database v5; SCREEN, search candidate cis-regulatory elements by the encyclopedia of DNA elements (ENCODE) database; SEdb, super-enhancer database.